Training load and baseline characteristics associated with new injury/pain within an endurance sporting population: A prospective study.

Journal:	*International Journal of Sports Physiology and Performance*
Manuscript ID	IJSPP.2018-0644.R1
Manuscript Type:	Original Investigation
Date Submitted by the Author:	28-Sep-2018

Complete List of Authors:	Johnston, Richard; University of Limerick Faculty of Education and Health Sciences, Physical Education and Sport Sciences Cahalan, Roisin; University of Limerick Faculty of Education and Health Sciences, School of Allied Health, University of Limerick, Ireland Bonnett, Laura; University of Liverpool, Department of Bio statistics Maguire, Matthew; Ulster Rugby, Irish Rugby Football Union, Strength and Conditioning Nevill, Alan; University of Wolverhampton, School of Sport, Performing Arts & Leisure Glasgow, Phil; Irish Rugby Football Union, Rehabilitation and Physiotherapy O'Sullivan, Kieran; Aspetar Qatar Orthopaedic and Sports Medicine Hospital, Sports Spine Centre Comyns, Tom; University of Limerick, PESS

Keywords:	training, endurance training, injury management

Training load and baseline characteristics associated with new injury/pain within an endurance sporting population: A prospective study.

Richard Johnston[1,4], Roisin Cahalan[2,8], Laura Bonnett[3] Matthew Maguire[4], Alan Nevill[5], Philip Glasgow[6], Kieran O'Sullivan[2,7,8], Thomas Comyns[1,8]

[1] Department of Physical Education and Sport Sciences, University of Limerick, Ireland

[2] School of Allied Health, University of Limerick, Ireland

[3] Department of Biostatistics, University of Liverpool, Merseyside, United Kingdom

[4] Ulster Rugby, Irish Rugby Football Union, Kingspan Stadium, Belfast, N.Ireland

[5] Faculty of Education Health and Wellbeing, University of Wolverhampton, England

[6] Irish Rugby Football Union, Lansdowne Road, Dublin 4, Ireland

[7] Sports Spine Centre, Aspetar Orthopaedic and Sports Medicine Hospital, Doha, Qatar

[8] Health Research Institute, University of Limerick, Ireland

Corresponding Author:

Richard Johnston

Department of Physical Education and Sport Sciences, University of Limerick, Ireland.

Tel: +447874006506

Fax: +447874006506

Email: Richard.Johnston@ul.ie

Word count: 2973

Abstract word count: 249

Number of tables: 3

Number of Figures: 3

Supplementary file: 2

Abstract

Purpose: To determine the association between training load (TL) factors, baseline characteristics and new injury and/or pain (IP) risk within an endurance sporting population (ESP).

Methods: Ninety-five endurance sporting participants from running, triathlon, swimming, cycling and rowing disciplines. Participants initially completed a questionnaire capturing baseline characteristics. TL and IP data was submitted weekly over a 52-week study period. Cumulative TL factors, Acute:Chronic Workload Ratios (ACWR) and Exponentially Weighted Moving Averages (EWMA) were calculated. A shared frailty model was used to explore time to new IP and association to TL factors and baseline characteristics.

Results: 92.6% of the ESP completed all 52 weeks of TL and IP data. The following factors were associated with the lowest risk of a new IP episode; (a) a low to moderate 7-day lag EWMA (0.8-1.3: HR=1.21, 95% CI 1.01-1.44, p=0.04) (b) a low to moderate 7-day lag weekly training load (WL) (1200-1700AU: HR=1.38, 95% CI=1.15-1.65, $p<0.001$) (c) a moderate to high 14-day lag 4-weekly cumulative training load (CL) (5200-8000AU: HR=0.33, 95% CI=0.21-0.50, $p<0.001$) and (d) a low number of previous IP episodes in the preceding 12 months (1 previous IP episode: HR=1.11, 95% CI=1.04-1.17, p=0.04).

Conclusions: To minimise new IP risk an ESP should avoid high spikes in acute TL whilst maintaining moderate to high chronic TLs. A history of previous IP should be considered when prescribing TLs. The demonstration of a lag between a TL factor and its impact on new IP risk may have important implications for future ESP TL analysis.

Keywords: workload, risk, single-discipline and multi-discipline sports

Introduction

Internationally endurance sports are growing in popularity with participants training and competing at recreational and elite level.[1] A unique characteristic of endurance sport populations (ESPs) is the heterogeneity of the training undertaken across different disciplines including running, cycling, triathlon, swimming and rowing. ESPs can be exposed to high training loads (TLs) and competition frequency which may contribute to the high prevalence of injury and/or pain (IP) (47-75%) reported within this population.[2] A surge in TL and IP research in recent years[3] has identified poor TL management as an IP risk factor. A consensus statement from the 2016 conference 'Monitoring Athlete Training Loads'[3] advises monitoring of both internal and external TLs. External TL is the objective physical load applied to the athlete[4] (e.g. distance covered, duration of session or frequency of sessions).[5] Internal TL is the individual physiological and/or psychological response to an external TL.[5] Several TL models can be derived from these internal and external TL measures, including Acute:Chronic Workload Ratios (ACWR) and Exponentially Weighted Moving Averages (EWMA).

ACWR capture the dynamic nature of training by allowing comparison of the acute TL undertaken, e.g. over 7 days, to the chronic TL undertaken, e.g. over 28 days.[6] Research within non-ESPs[7] has identified ACWR parameters, or 'sweet spots', which are associated with lower relative IP risk in soccer (ACWR 1.0-1.25),[8] rugby league (ACWR 0.85-1.35)[9] and cricket (0.8-1.3).[10] The training-injury prevention paradox model[7] proposes that such 'sweet spot' ACWR balance the potential positive effects of chronic TL (e.g. fitness) with the potential negative effects of high 'spikes' in acute TL (e.g. fatigue). However, more recent non-ESP studies have moved towards reporting ACWR using an EWMA method.[11,12] EWMA addresses the decaying nature of fitness and the non-linear nature of TL.[13] It assigns a decreasing weighting to each older TL thereby giving more weighting to recent acute TLs and less weighting to previous chronic TLs.[11] However, both ACWR and EWMA have not yet been investigated and characterised within ESPs.[5,6]

Increasingly within non-ESP research there has been a shift from reporting traditional TL factors in isolation and an increased appreciation of the complex relationship between TL, athlete baseline characteristics and IP risk.[14] A recent systematic review of ESPs[6] has identified non-modifiable baseline characteristics (i.e. increased age, history of previous IP) which are associated with increased IP risk. The aim of this prospective study was to determine the association between TL factors, including ACWR/EWMA, and new IP risk within an ESP. The study also aimed to further define the association between non-modifiable baseline characteristics and new IP risk within an ESP.

Methods

Subjects

116 participants were initially recruited from 15 ESP clubs and, other than age (18-65 years), no exclusion criteria were applied. Both elite (5%) and recreational participants (95%) were included in the study population. Elite or recreational level was self-reported by the participants. Ethical approval was granted by a local university and informed and written consent was provided by all participants.

Methodology

Participants completed a baseline questionnaire relating to non-modifiable baseline characteristics (e.g. age, sex, history of IP), as well as training profile and endurance sporting experience. A preliminary pilot test of the questionnaire and electronic training diary was completed prior to study commencement. Over the following 52 weeks the participants utilised an electronic SurveyMonkey™ online 'training diary' to upload TL and IP data weekly. Participants received an email with a link to the diary on the Sunday of each week and an email reminder four days later from the lead author (RJ). The questionnaire asked participants to report validated training data on; (1) each training/competition event, (2) day of the week, (3) session type (e.g. running, swimming), (4) duration (minutes), (5) distance (meters/kilometers) and (6) intensity (session training load (sRPE))[15] (Borg CR-10 scale).[15] Participants also subjectively recorded any IP episode by body location each week. 21 participants were removed due to submitting insufficient training data (<30 weeks) resulting in a final study population of 95 participants across five endurance disciplines. (Table 1).

Based on recommendations from the International Olympic Committee[16] (IOC) and a previous editorial,[17] an IP episode was defined as any physical musculoskeletal complaint/impairment, solely due to participation in endurance discipline training and/or competition event, which may have caused the participant to continue to train/compete fully or reduce/adapt or miss time from training/competition. This definition was provided to participants in the electronic training diary. If a participant reported an initial IP episode in a particular body location it was categorised as a new IP episode. If the participant, then reported an IP episode in the same body location in the subsequent four weeks it was categorised as a secondary IP episode.[18]

Data was categorized into weekly blocks (1-52) running from each Monday to Sunday. If a participant did not perform a daily/weekly TL, a value of zero was included to allow analysis of new IP risk following no TL.[19] To identify if a TL factor contributed to the onset of a new IP episode a 7 and 14-day time lag was implemented.[10,20,21] That is, if a new IP episode was reported in week 10 then TL was analysed for week 9 (i.e. 7-day lag) and week 8 (i.e. 14-day lag).[22] TL factors (Table 2) were calculated using Microsoft Excel software (2016).

Statistical analysis

TL and baseline characteristic variables were summarised according to total and percentage differences between types of endurance discipline, assessed via chi-squared tests or Fisher's exact tests in the case of small responses in at least one category (Table 1). Chi-squared tests and Fisher's exact tests summarised normally distributed data as mean and standard deviations and skewed continuous data as median and interquartile ranges. New IP rates were expressed as the total number of new IP/total number of training sessions performed and reported per 1000 training sessions.[12] Missing data (<5% for each variable) was imputed with the median response for that variable.[23]

A shared frailty model was used to estimate times between new IP episodes. The model used random effects following a gamma distribution, with a mean equal to one and unknown variance to account for the within participant correlation between new IP episodes. A restricted maximum likelihood criterion was used to choose the variance of the random effect. Results were presented as Hazard Ratios (HRs) with 95% confidence intervals and a p value (≤0.05) indicating results of statistical significance. This model was adopted due to its use in sports medicine[22] literature and as it allows multiple IP episodes for each participant to be analysed as the outcome of interest.[24] A parsimonious model was built from a pool of 18 variables via backwards selection according to Akaike's Information Criterion.[25] Results for the continuous variables were presented as post-hoc defined categorical variables, with categories chosen according to knot positions for a spline model fit to the data. The TL categories (Table 3) were derived from previous ESP studies[26] and non-ESP studies[8,14,21,27,28] and adapted to ensure approximately even distribution of the TL data across the categories. In line with all previous non-ESP studies[8,14,21,27,28] the lowest range was assigned as the reference range.

Discrimination of the model was assessed using the c-statistic which differentiated between those who reported IP and those who did not. The c-statistic is equivalent to the area under the Receiver Operator Characteristic (ROC) curve and is measured on a scale ranging from 0.5 (no better than chance) to 1 (perfect prognostic). The c-statistic for this IP modelling was 0.70 (0.65 to 0.73) a good fit overall. Analyses were performed using R version 3.2.3 using the 'survival' package. Computer code for all analyses, including the list of standard packages used as part of the analysis, are available in supplementary file 1.

Results

89 of the 95 participants (92.6%) submitted TL and IP data for all weeks of data collection. Table 1 displays the median values for each TL factor across the study period. The mean prevalence of new IP was 6.1 per participant with a rate of new IP 0.12 per session. Within endurance athlete subgroups, runners accounted for over half (53.1%) of new IP episodes. The lower limb (foot, shin/calf) accounted for 20.1% of new IP episodes (supplementary table 1).

Ten out of eighteen prognostic variables (14-day lag WL, 7-day lag CL, 7 and 14-day lag training strain and training monotony, W-WL, 7 and 14-day lag ACWR and 14-day lag EWMA), analysed in the parsimonious multivariable model, did not reach statistical significance. Eight prognostic variables reached or were close to statistical significance ($p<0.05$) (Table 3).

IP and TLs (Table 3)

7-day lag WL (Figure 1) demonstrated a positive linear effect, with increasing WL significantly associated with increasing new IP risk (HR=1.46, CI 95%=1.18-1.81, $p<0.001$). Whilst 7-day lag CL was not found to be significantly associated, a non-linear effect was evident with the 14-day lag CL (Figure 2). Whilst a low 14-day lag CL (2000-3500AU) was associated with reduced new IP risk (HR 0.73, 95% CI 0.65-0.82, $p<0.001$), a moderate to high 14-day lag CL (3500-5200AU and 5200-800AU) was associated with a greater reduction in new IP risk (HR=0.47, CI 95%=0.36-0.63, $p<0.001$ and HR=0.33, CI 95%=0.21-0.50, $p<0.001$). However, very high 14-day lag CL (>8000AU) increased the risk of a new IP episode (HR=1.71, CI 95%=2.09-1.40, $p<0.001$).

The lowest risk of new IP was demonstrated with a low to moderate 7-day lag EWMA of 0.8-1.3 (HR 1.21, 95% CI 1.01-1.44, p=0.04), when compared to the reference range of <0.8 (Figure 3). As the 7-day lag EWMA increases the risk of new IP increases, with very high EWMA (>1.5) associated with the highest new IP risk (HR 2.15, 95% CI 1.04-4.44, p=0.04). There was no association between 14-day lag EWMA and new IP risk (p>0.05). No significant association between new IP risk and the number of training sessions per week was demonstrated (p=0.06).

IP and baseline characteristics

Reporting IP in the previous 12 months was associated with new IP risk (HR = 1.11, CI 95% = 1.01-1.21, p=0.04). As the number of previous IP episodes reported increased the risk of new IP also increased, the highest risk associated with ≥3 previous IP episodes (HR=1.92, 95% CI 1.31-2.81, p=0.04). Age (p=0.58) and sex (p=0.14) were not associated with increased risk of new IP.

Discussion

This prospective study is the first to investigate the association between internal and cumulative TL factors and new IP risk within an ESP. Excellent completion and retention rates were observed, with 92.6% participants completing all 52 weeks of TL and IP data. The results demonstrate that a low to moderate 7-day lag WL, a moderate to high 14-day lag CL and a low to moderate 7-day lag EWMA are associated with the lowest risk of a new IP episode within an ESP. These results support the IOC consensus statement[16] which highlights the importance of utilising internal and cumulative TL factors in the identification of IP risk. A history of previous IP was found to be significantly associated with new IP risk whilst ESP sex and age were not associated with new IP risk.

Training Load

Previous studies[8,10,12] within non-ESPs have identified that a 'sweet spot' ACWR of 0.8-1.3 is associated with lower IP risk. However, an association between 7-day and 14-day lag ACWR and new IP risk was not identified within this ESP study. A potential reason for this is that ACWR utilises rolling averages to assign the same relative weight to all TLs in both the acute and chronic training windows. However recent research suggests recently accumulated training has a greater impact on fitness and fatigue than the preceding weeks of training.[7,24] There are also inherent differences between ESPs and non-ESPs with approximately 80% of ESP TL conducted at lower intensities[29] whilst non-ESP TL favours moderate to high training intensities.[6] The overall median 7-day lag WL within this ESP was lower (1130AU) than previous non-ESP studies,[14,28] therefore the ACWR may not be sensitive in detecting subtle changes in acute TL within an ESP. EWMA, a derivative of ACWR, applies a decaying function to give a greater weight to recently-completed TL. Two recent non-ESP[11,12] studies both found EWMA to provide a more sensitive IP risk model when compared to ACWR.

7-day lag EWMA did demonstrate an association with new IP risk within the ESP, that is as the 7-day lag EWMA increased the risk of new IP increased (Figure 3), with a 2-fold increase in new IP episodes with very high spikes in 7-day lag EWMA (>1.5). This finding is in line with previous non-ESP research which concur that excessive and rapid spikes in acute TL increase IP risk.[4,8,9,12,16] Research in team-based sports[11,12,20] has shown that if acute load is too high (e.g. high levels of fatigue) and chronic load is too low (e.g. low levels of fitness) then the athlete will be in a more fatigued state with increased IP risk.[7] Inversely, if the acute training load is lower (e.g. the athlete is experiencing minimal fatigue) and the chronic load is higher (e.g. the athlete is developing fitness) then the athlete is in a well prepared state with a low IP risk.[7] This is supported in this ESP where low and moderate 7-day lag WLs (1200-1700AU and 1700-2200AU) are associated with a lower new IP risk than high and very high 7-day lag WLs (2200-2700AU and >2700 AU) (Figure 1). Previous non-ESP studies have also demonstrated that high spikes in 7-day lag WL (>1245AU,[28] >1500AU[8] and >1750AU[14]) are associated with increased IP risk.

It is interesting to note 14-day lag WL was not associated with new IP risk in this ESP, suggesting that the negative consequences of fatigue manifest in the week following an acute high spike in TL. The concept of a potential lag between a TL and its positive (e. g. fitness, strength, robustness) and negative (e.g. fatigue) consequences has been described in non-ESP research.[9,20] This is supported by the finding in the current study that a 14-day lag CL was associated with new IP risk, whilst a 7-day lag CL was not. This suggests that the beneficial effects of CL (e.g. fitness) does not manifest in the acute period but rather after a 14-day lag. A moderate to high 14-day lag CL (3500-8000AU) was associated with the lowest new IP risk whilst both a low (2000-3500AU) and very high (>8000AU) 14-day lag CL were associated with higher new IP risk (Figure 2). This finding reflects Gabbett's[7] proposed training-injury prevention paradox whereby a minimum TL is required to produce beneficial training adaptations over time and protect against IP. Low 14-day lag CLs are unlikely to be sufficient to maintain fitness and allow adaptations, whilst striving to maintain very high 14-day lag CLs is likely to result in fatigue. Thus, to minimise the risk of a new IP episode the ESP should maintain a low to moderate acute TL (i.e. 7-day lag EWMA 0.8-1.3, 7-day lag 1200-1700AU) to protect against fatigue, whilst ensuring chronic TLs are sufficient to develop and maintain fitness (i.e. 14-day lag CL 3500-8000AU).

Whilst an association between frequency of training sessions per week and new IP risk did not meet significance (p=0.06), the trend of increased new IP risk with high training frequency (≥5 sessions/week) is in keeping with previous reviews in non-ESPs.[30] A high frequency of training sessions suggests insufficient recovery periods between training sessions, increasing the risk of fatigue. Both training monotony and training strain were not found to be significantly associated with increased IP risk within this ESP. To the authors knowledge training monotony and strain have not been previously studied in ESPs. One potential reason for the low level of training monotony (0.77) within this ESP is the low frequency of training sessions (median=4 per week) conducted by the ESP compared to a non-ESP study[19] (median=8 sessions per week).

Baseline characteristics

As previously reported in ESP systematic reviews[5,6] a history of previous IP was associated with increased new IP risk. In particular the greater the number of previous IP episodes, the greater the risk of a new IP episode. This may reflect not only the negative impact of previous IP on current TLs and fitness, but also a pattern of TL mismanagement or athlete frailty. Increasing age did not demonstrate a significant association with new IP risk (p=0.58).

Limitations

New IP episodes were reported subjectively and are therefore open to reporting bias. There was heterogeneity across the study population with large differences in the sporting disciplines represented, with 59% of the study population runners and only 2.1% swimmers. Whilst both elite and recreational participants were included in the study, no definition was provided to participants when they were asked to subjectively report whether they were at an elite or recreational level. Whilst TL and IP data was collected over a 52-week period no statistical analysis was conducted in relation to training and competition blocks.

Practical Applications

To minimise the risk of new IP an ESP should maintain low to moderate acute TLs and avoid high spikes in acute TL. ESPs should also aim to gradually develop and maintain fitness through moderate to high chronic TLs. Clinical practice within ESPs should implement the routine use of cumulative TL measures, in particular EWMA which may be a more sensitive acute TL model in ESPs.

Conclusions

This study is the first to characterise associations between TL factors and IP risk within an ESP. The lowest risk of a new IP episode was observed when the acute TL was low to moderate and the chronic TL was moderate to high. This study also highlights a potential lag between a TL and its subsequent impact upon new IP risk. As a history of previous IP was associated with increased new IP risk, this should also be considered when prescribing TLs within an ESP.

Acknowledgments

The authors would like to acknowledge with considerable gratitude the participants for recording training load and injury and/or pain data throughout the study period. The faculty of Education and Health Sciences at the University of Limerick funds Richard Johnston through a PhD scholarship.

REFERENCES

1. Visentini P, Clarsen B. Oversue Injuries In Cycling: The Wheel Is Turning Towards Evidence-Based Practice. *Aspetar Sports Med J.* 2017;6:486-492.
2. Andersen CA, Clarsen B, Johansen TV, Engebretsen L. High prevalence of overuse injury among iron-distance triathletes. *Br J Sports Med.* 2013;47(13):857-861.
3. Bourdon PC, Cardinale M, Murray A, et al. Monitoring Athlete Training Loads: Consensus Statement. *Int J Sports Physiol Perform.* 2017;12(Suppl 2):S2-161-S162-170.
4. Gabbett TJ. Reductions in pre-season training loads reduce training injury rates in rugby league players. *Br J Sports Med.* 2004;38(6):743.
5. Hulme A, Nielsen RO, Timpka T, Verhagen E, Finch C. Risk and Protective Factors for Middle- and Long-Distance Running-Related Injury. *Sports Med.* 2017;47(5):869-886.
6. Johnston R, Cahalan R, O'Keeffe M, O'Sullivan K, Comyns T. The associations between training load and baseline characteristics on musculoskeletal injury and pain in endurance sport populations: A systematic review. *J Sci Med Sport.* 2018.
7. Gabbett TJ. The training-injury prevention paradox: should athletes be training smarter and harder? *Br J Sports Med.* 2016;50(5):273-280.
8. Malone S, Owen A, Newton M, Mendes B, Collins KD, Gabbett TJ. The acute:chonic workload ratio in relation to injury risk in professional soccer. *J Sci Med Sport.* 2017;20(6):561-565.
9. Hulin BT, Gabbett TJ, Lawson DW, Caputi P, Sampson JA. The acute:chronic workload ratio predicts injury: high chronic workload may decrease injury risk in elite rugby league players. *Br J Sports Med.* 2016;50(4):231-236.
10. Hulin BT, Gabbett TJ, Blanch P, Chapman P, Bailey D, Orchard JW. Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *Br J Sports Med.* 2014;48(8):708-712.
11. Murray NB, Gabbett TJ, Townshend AD, Blanch P. Calculating acute:chronic workload ratios using exponentially weighted moving averages provides a more sensitive indicator of injury likelihood than rolling averages. *Br J Sports Med.* 2017;51(9):749.
12. Sampson JA, Murray A, Williams S, et al. Injury risk-workload associations in NCAA American college football. *J Sci Med Sport.* 2018.
13. Williams S, West S, Cross MJ, Stokes KA. Better way to determine the acute:chronic workload ratio? *Br J Sports Med.* 2017;51(3):209.
14. Rogalski B, Dawson B, Heasman J, Gabbett TJ. Training and game loads and injury risk in elite Australian footballers. *J Sci Med Sport.* 2013;16(6):499-503.
15. Foster C, Florhaug JA, Franklin J, et al. A new approach to monitoring exercise training. *J Strength Cond Res.* 2001;15(1):109-115.
16. Soligard T, Schwellnus M, Alonso J-M, et al. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med.* 2016;50(17):1030-1041.

17. O'Sullivan K, O'Sullivan PB, Gabbett TJ. Pain and fatigue in sport: are they so different? *Br J Sports Med.* 2018;52(9):555.
18. Møller M, Nielsen RO, Attermann J, et al. Handball load and shoulder injury rate: a 31-week cohort study of 679 elite youth handball players. *Br J Sports Med.* 2017;51(4):231.
19. Comyns T, Flanagan EP. Applications of the Session Rating of Perceived Exertion System in Professional Rugby Union. *Strength Cond Jl.* 2013;35(6):78-85.
20. Carey DL, Blanch P, Ong K-L, Crossley KM, Crow J, Morris ME. Training loads and injury risk in Australian football—differing acute: chronic workload ratios influence match injury risk. *Br J Sports Med.* 2017;51:1215-1220.
21. Piggott B, Newton, M. J., & McGuigan, M. R. The relationship between training load and incidence of injury and illness over a pre-season at an Australian football league club. *J Austr Strength Cond Res.* 2009;17(3):4-17.
22. von Rosen P, Frohm A, Kottorp A, Friden C, Heijne A. Multiple factors explain injury risk in adolescent elite athletes: Applying a biopsychosocial perspective. *Scand J Med Sci Sports.* 2017.
23. Little RJA, Rubin DB. *Statistical analysis with missing data.* John Wiley & Sons, Inc.; 1986.
24. Windt J, Gabbett TJ. How do training and competition workloads relate to injury? The workload-injury aetiology model. *Br J Sports Med.* 2017;51:428-435.
25. Akaike H. Akaike's Information Criterion. In: Lovric M, ed. *International Encyclopedia of Statistical Science.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2011:25-25.
26. Malisoux L, Nielsen RO, Urhausen A, Theisen D. A step towards understanding the mechanisms of running-related injuries. *J Sci Med Sport.* 2015;18(5):523-528.
27. Malone S, Roe M, Doran DA, Gabbett TJ, Collins KD. Protection Against Spikes in Workload With Aerobic Fitness and Playing Experience: The Role of the Acute:Chronic Workload Ratio on Injury Risk in Elite Gaelic Football. *Int J Sports Physiol Perform.* 2017;12(3):393-401.
28. Cross MJ, Williams S, Trewartha G, Kemp SP, Stokes KA. The Influence of In-Season Training Loads on Injury Risk in Professional Rugby Union. *Int J Sports Physiol Perform.* 2016;11(3):350-355.
29. Seiler S. What is best practice for training intensity and duration distribution in endurance athletes? *Int J Sports Physiol Perform.* 2010;5(3):276-291.
30. Jones CM, Griffiths PC, Mellalieu SD. Training Load and Fatigue Marker Associations with Injury and Illness: A Systematic Review of Longitudinal Studies. *Sports Med.* 2016.

Table 1: Endurance population characteristics

Variable	Total population N=95	Runners N= 56 (59.0%)	Triathletes N=18 (18.9%)	Swimmers N=2 (2.1%)	Cyclists N=10 (10.5%)	Rowers N=9 (9.5%)	P value
Males, N (%)	61 (64.2%)	29 (47.6%)	16 (26.2%)	1 (1.6%)	9 (14.8%)	6 (9.8%)	**0.02***
Age (yrs), mean ±SD	42.2 ± 10.0	42.3 ± 8.8	40.2 ± 7.4	34.5 ± 20.5	42.1 ± 11.3	48.1 ±16.5	0.30
Total new IP, N (%)	585 (100)	311 (53.1)	140 (23.7)	17 (2.9)	70 (11.8)	50(8.5)	**<0.001***
WL (AU) (IQR)	1130 (630:1740)	1005 (530:1599)	1465 (870:2160)	1890 (360:4905)	1225 (783:1735)	1070 (690:1520)	**<0.001***
CL (AU) (IQR)	4370 (2550:6405)	3930 (2070:5915)	5498 (3520:7985)	9303 (840:17749)	4800 (3465:6311)	4235 (3045:5720)	**<0.001***
W-WL (AU) mean ± SD	0.25 ± 897	1.90 ± 851	-1.68 ± 867	-43.3 ± 1318	-1.45 ± 1090	5.71 ± 867	0.96
Training monotony (IQR)	0.77 (0.59:1.01)	0.77 (0.59:1.01)	0.88 (0.68:1.14)	0.76 (0.59:1.06)	0.63 (0.52:0.78)	0.76 (0.61:0.95)	**<0.001***
Training strain (IQR)	895 (411:1437)	809 (340:1498)	1309 (666:2338)	1367 (196:5202)	828 (447:1254)	830 (450:1384)	0.19
ACWR rolling average (7:28 days) (IQR)	1.02 (0.78:1.26)	1.02 (0.78:1.26)	1.01 (0.82:1.23)	1.05 (0.87:1.33)	1.04 (0.69:1.33)	1.05 (0.76:1.29)	0.81
EWMA moving average 7:28 days	1.00 (0.84:1.18)	1.00 (0.83:1.19)	1.00 (0.84:1.14)	1.01 (0.78:1.24)	1.00 (0.84:1.18)	1.03 (0.91:1.20)	0.15

N=Number; p=power; yrs=years; SD=Standard Deviation; AU=Arbitrary Unit; wk= weekly; WL=weekly training load; CL=4-weekly cumulative training load; W-WL= Week-to-week change in training load; IQR=Inter Quartile Range; ACWR=Acute:Chronic Workload Ratio; EMWA=Exponentially Weighted Moving Average.

1 **Table 2: Training load factor definitions and calculations**

Training load factor	**Definition**	**Calculation**
Session training load (sRPE)[15,19]	Measure of session internal and external training load	sRPE = Session duration (mins) x session intensity (Borg CR-10 scale)
Daily training load (DL)[15,19]	Measure of daily training load	DL = Sum of all sRPE for that day
Weekly training load (WL)[12,18,27]	Measure of weekly training load	WL = Sum of all DLs
7-day lag WL	A measure of the WL 7 days before a new IP episode	
14-day lag WL	A measure of the WL 14 days before a new IP episode	
Four weekly cumulative training load (CL)[12,18,27]	Measure of cumulative four-weekly training loads	CL = Sum of all sRPE per four weeks
7-day lag CL	A measure of the CL 7 days before a new IP episode	
14-day lag CL	A measure of the CL 14 days before a new IP episode	
Week-to-week change in training load (W-WL)[12,18,27]	Absolute difference between current and previous week's training load	W-WL = Current week WL – previous week WL
Training monotony[15,19]	A measure of day-to-day training variability during a training week.	Monotony = Mean DL ÷ standard deviation of DL over 1wk
Training strain[15,19]	A measure which represents the overall stress that an athlete was exposed to throughout the training week.	Training strain = WL x Training monotony
(Coupled) Acute:chronic workload (ACWR) with 7 and 14 day lag [13,14,24]	Calculated by expressing a rolling average of an athlete's training load completed in an acute period (seven days) with the chronic training load completed over a longer period (twenty-eight days)	ACWR = current WL ÷ (previous mean CL)

Exponentially weighted moving average (EWMA) with a 7 and 14-day lag[11,15,18]	Calculated by expressing a moving average of an athlete's training load completed in an acute period (seven days) and chronic period (twenty-eight days). This method assigns a decreasing weighting to compensate for the latency effects of training loads.	$EWMA_{week} = WL \times \lambda a + ((1- \lambda a) \times EWMA_{28day}$

λa= a value between 0 and 1 that represents the degree of training decay

Table 3: Injury and/or pain data

Variable	**Comparison**	**p-value**	**HR (95% CI) - continuous**	**HR (95% CI) – post-hoc categorisation**
Endurance athlete subgroup	Runner (ref)		1.00	
	Triathlete	0.09	1.39 (0.95-2.07)	
	Swimmer	0.65	1.25 (0.47-3.35)	
	Cyclist	**0.02***	**1.76 (1.11-2.79)**	
	Rower	0.61	1.14 (0.69-1.88)	
Sex	Male (ref)	-	1.00	
	Female	0.14	1.25 (0.93-1.69)	
Age	<27(ref)	0.58	1.04 (0.90-1.21)	1.00
	27 to 36			1.03 (0.95-1.12)
	36 to 43			1.07 (0.91-1.25)
	43 to 49			1.10 (0.88-1.37)
	49 to 59			1.13 (0.84-1.54)
	>59			1.17 (0.80-1.71)
History of IP	0 (ref)	**0.04***	**1.11 (1.01-1.21)**	**1.00**
	1			**1.11 (1.04-1.17)**
	2			**1.22 (1.09-1.37)**
	≥3			**1.92 (1.31-2.81**
Number of training sessions per week	0-3	0.06	1.05 (1.00-1.10)	0.91 (1.00-0.83)

	4 (ref)			1.00
	≥5			1.36 (1.01-1.83)
7-day lag WL (per 1000 AU)	0-1200 (ref)	**<0.001***	**1.46 (1.18-1.81)**	**1.00**
	1200-1700			**1.38 (1.15-1.65)**
	1700-2200			**1.67 (1.25-2.22)**
	2200-2700			**2.02 (1.36-2.98)**
	>2700			**4.14 (1.88-9.15)**
14-day lag CL (per 1000 AU)	0-2000 (ref)	**<0.001***	**0.82 (0.76-0.89)**	**1.00**
	2000-3500			**0.73 (0.65-0.82)**
	3500-5200			**0.47 (0.36-0.63)**
	5200-8000			**0.33 (0.21-0.50)**
	>8000			**1.71 (2.09-1.40)**
7-day lag EWMA (7:28 days) (per 0.1)	<0.8 (ref)	**0.04***	**1.03 (1.01-1.06)**	**1.00**
	0.8-1.3			**1.21 (1.01-1.44)**
	1.3 to 1.5			**1.34 (1.01-1.76)**
	>1.5			**2.15 (1.04-4.44)**

P= power; CI=confidence interval; AU=Arbitrary Unit; *=significant result; IP=Injury and/or Pain; Ref= reference range; WL= Weekly training load; CL= Four weekly cumulative training load; EWMA= Exponentially weighted moving average

Figure 1: Plot of the relative hazard ratio of new Injury and/or Pain against 7-day lag Weekly Training Load.

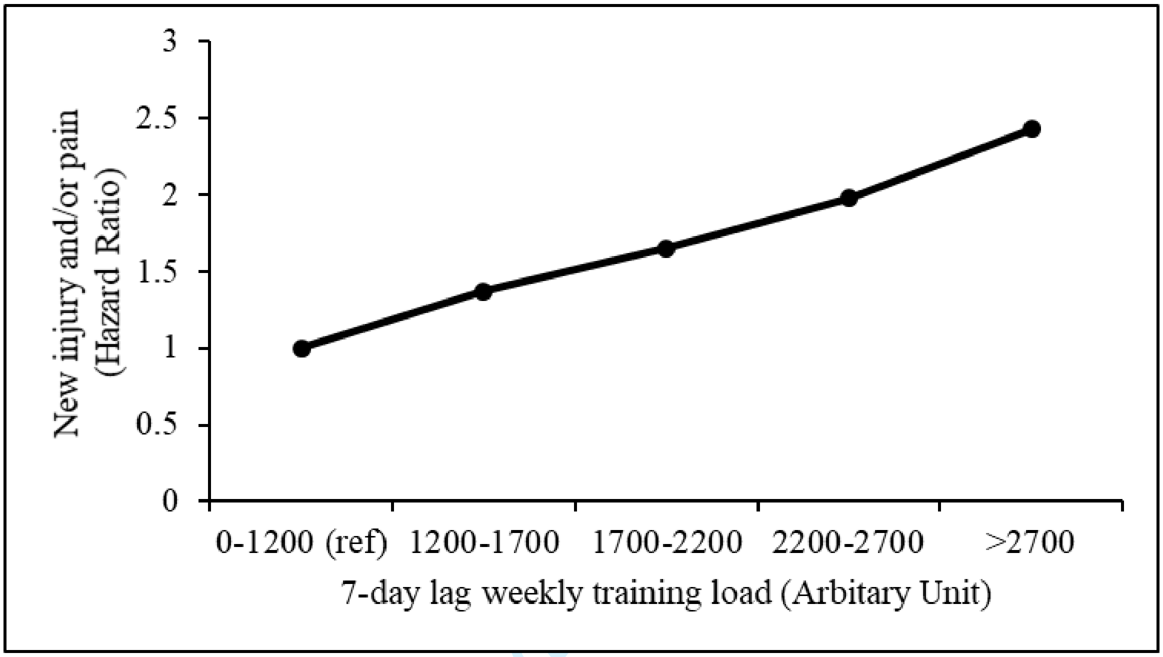

Figure 2: Plot of the relative hazard ratio of new Injury and/or Pain against 14-day lag Four Weekly Cumulative Training Load.

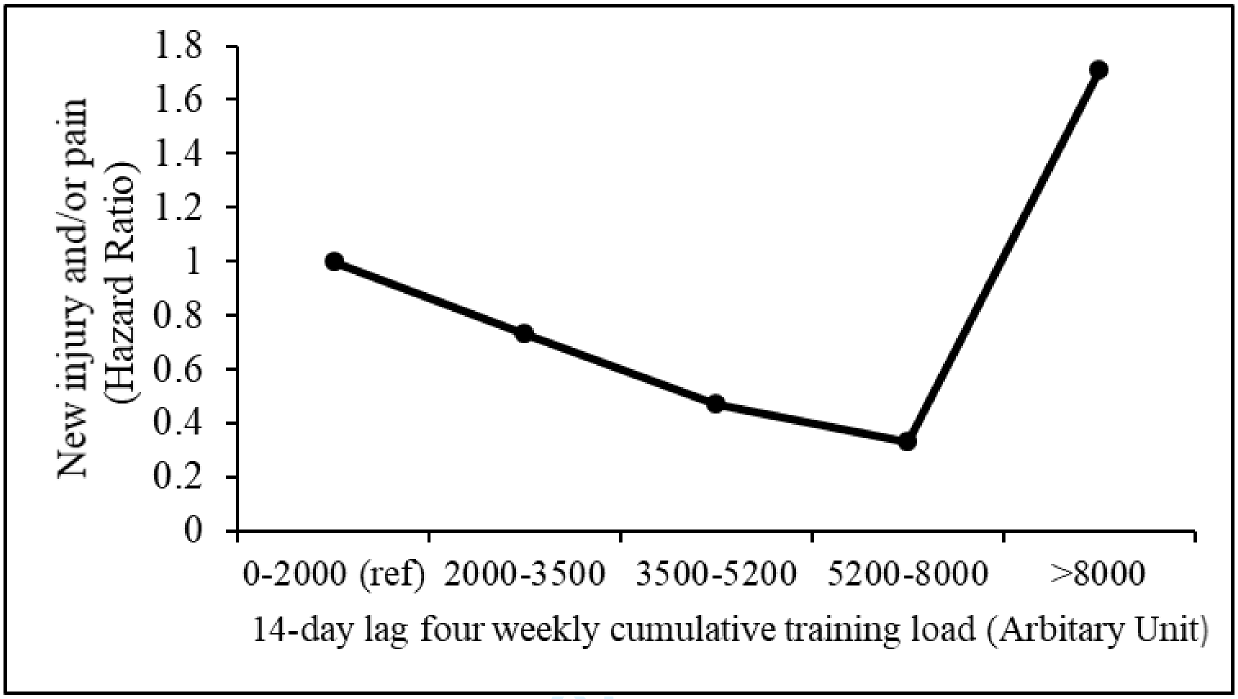

Figure 3: Plot of the relative hazard ratio of new Injury and/or Pain against 7-day lag Exponentially Weighted Moving Average.

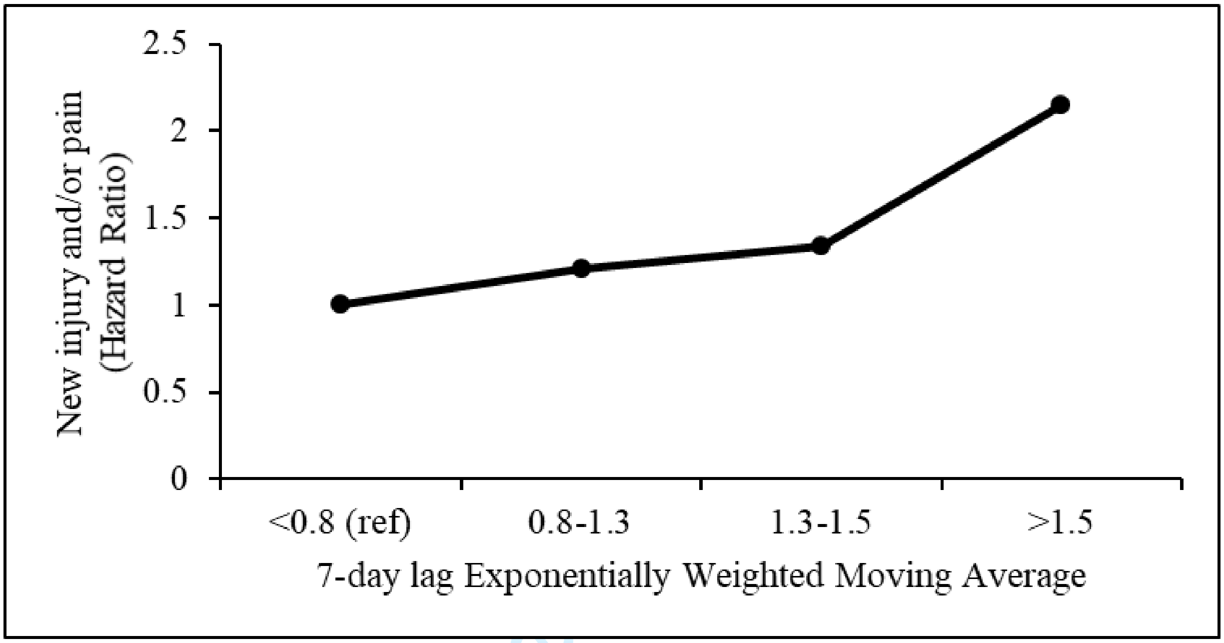

Figure captions

Figure 1: Plot of the relative hazard ratio of new Injury and/or Pain against 7-day lag Weekly Training Load.

Figure 2: Plot of the relative hazard ratio of new Injury and/or Pain against 14-day lag Four Weekly Cumulative Training Load.

Figure 3: Plot of the relative hazard ratio of new Injury and/or Pain against 7-day lag Exponentially Weighted Moving Average.

Frailty Model Code - Johnston et al

Written by: Dr Laura Bonnett

Load Packages

```
library(survival)
library(MASS)
library(plyr)
library(Hmisc)

## Loading required package: lattice

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:plyr':
##
##     is.discrete, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units

library(mfp)
library(MASS)
library(Hmisc)

richard <- read.csv("Combined dataset.csv", header=TRUE)
colnames(richard)[1] <- "Pre_lag_7_28_2wk"

# remove rows without ID
richard2 <- subset(richard,ID_12>0)
```

Frailty model - total injuries

```
# Replace missing injuries with zero
for(i in 1:nrow(richard2)){if(is.na(richard2$Total_IP)[i]) richard2$Total_IP[i] <- 0}

# Recode outcome so no events = 0 & injuries = 1 (not 1: all injuries; 2: no injuries))
for(i in 1:nrow(richard2)){if(richard2$Total_IP[i]>1) richard2$Total_IP[i] <- 1}

# Replace missing values with median
```

```
for(i in 1:nrow(richard2)){if(is.na(richard2$Total_TL)[i]) richard2$Total_TL[i] <- median(richard2$Total_TL,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$Abs_change)[i]) richard2$Abs_change[i] <- median(richard2$Abs_change,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$Pre_lag_7_28_2wk)[i]) richard2$Pre_lag_7_28_2wk[i] <- median(richard2$Pre_lag_7_28_2wk,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$Pre_lag_7_28)[i]) richard2$Pre_lag_7_28[i] <- median(richard2$Pre_lag_7_28,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$EWMA_pre_lag_2wk)[i]) richard2$EWMA_pre_lag_2wk[i] <- median(richard2$EWMA_pre_lag_2wk,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$EWMA_pre_lag_1wk)[i]) richard2$EWMA_pre_lag_1wk[i] <- median(richard2$EWMA_pre_lag_1wk,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$Total_TL_perwk_pre_2wk)[i]) richard2$Total_TL_perwk_pre_2wk[i] <- median(richard2$Total_TL_perwk_pre_2wk,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$Pre_2_Cum1_4)[i]) richard2$Pre_2_Cum1_4[i] <- median(richard2$Pre_2_Cum1_4,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$Pre_1_Cum1_2)[i]) richard2$Pre_1_Cum1_2[i] <- median(richard2$Pre_1_Cum1_2,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$Mon_pre_2wk)[i]) richard2$Mon_pre_2wk[i] <- median(richard2$Mon_pre_2wk,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$Mon_pre_1wk)[i]) richard2$Mon_pre_1wk[i] <- median(richard2$Mon_pre_1wk,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$train_strain_2wk)[i]) richard2$train_strain_2wk[i] <- median(richard2$train_strain_2wk,na.rm=TRUE)}
for(i in 1:nrow(richard2)){if(is.na(richard2$train_strain_1wk)[i]) richard2$train_strain_1wk[i] <- median(richard2$train_strain_1wk,na.rm=TRUE)}
```

Frailty Model – adjusted variables

```
ts2wk100 <- richard2$train_strain_2wk/100
ts1wk100 <- richard2$train_strain_1wk/100

TL1000 <- richard2$Total_TL/1000
TLpre2wk1000 <- richard2$Total_TL_perwk_pre_2wk/1000

precum14 <- richard2$Pre_2_Cum1_4/1000
precum12 <- richard2$Pre_1_Cum1_2/1000

EWMApre2wk <- richard2$EWMA_pre_lag_2wk*10
EWMApre1wk <- richard2$EWMA_pre_lag_1wk*10

Abs_change1000 <- richard2$Abs_change/1000

age <- richard2$Age_12/10
```

Frailty Model - backwards elimination

```
fitf <- coxph(Surv(week_num,Total_IP)~factor(END_sub)+SEX+age+Num_sessions+TL1000+Abs_change1000+Pre_lag_7_28_2wk+Pre_lag_7_28+Previous_injury+EWMApre2wk+EWMApre1wk+TLpre2wk1000+precum14+precum14+Mon_pre_2wk+Mon_pre_1wk+ts2wk100+ts1wk100+frailty(ID_12,dist="gaussian"),data=richard2)

backward_mod <- stepAIC(fitf,scope=list(upper=~factor(END_sub)+SEX+age+Num_sessions+TL1000+Abs_change1000+Pre_lag_7_28_2wk+Pre_lag_7_28+Previous_injury+EWMApre2wk+EWMApre1wk+TLpre2wk1000+precum14+precum14+Mon_pre_2wk+Mon_pre_1wk+ts2wk100+ts1wk100+frailty(ID_12,dist="gaussian"),lower=~frailty(ID_12,dist="gaussian")),direction="backward",trace=FALSE)

summary(backward_mod)
```

1 **Supplementary table 1: Location of new injury and/or pain**

Area of IP	Total population N=95	Runners N= 56 (59.0%)	Triathletes N=18 (18.9%)	Swimmers N=2 (2.1%)	Cyclists N=10 (10.5%)	Rowers N=9 (9.5%)	p value
Upper leg (quadriceps/hamstring)	53 (100.0%)	33 (62.2%)	10 (18.8%)	0 (0.0)	6 (11.5%)	4 (7.5%)	**<0.001***
Lower limb (foot, shin/calf)	124 (100.0%)	72 (58.0%)	37 (29.8%)	4 (3.2%)	7 (5.8%)	4 (3.2%)	**<0.001***
Knee	73 (100.0%)	36 (49.3%)	18 (24.6%)	1 (1.3%)	14(19.1%)	4 (5.7%)	**<0.001***
Hip/groin	92 (100.0%)	59 (64.1%)	18 (19.5%)	1 (1.0%)	5 (5.6%)	9 (9.8%)	**<0.001***
Pelvis	2 (100.0%)	2 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	**<0.01***
Lower back	112 (100.0%)	52 (46.4%)	28 (25%)	1 (1.1%)	16 (14.2%)	15 (13.3%)	**<0.001***
Upper back	16 (100.0%)	6 (37.5%)	3 (18.7%)	1 (6.2%)	5 (31.4%)	1 (6.2%)	0.08
Neck	8 (100.0%)	3 (37.5%)	2 (25.0%)	0 (0.0%)	3 (37.5%)	0 (0.0%)	0.06
Chest	10 (100.0%)	5 (50.0%)	2 (20.0%)	1 (10.0%)	1 (10.0%)	1 (10.0%)	0.06
Upper limb	95 (100.0%)	43 (45.2%)	22 (23.3%)	8 (8.4%)	10 (10.5%)	12 (12.6%)	**<0.001***

2 IP = Injury and/or Pain; N = Number; p = power